FILOZOFICKÁ FAKULTA
UNIVERZITY KARLOVY
V PRAZE

Ústav anglického jazyka a didaktiky

# BAKALÁŘSKÁ PRÁCE

Dita Fejlová

**Relationship between the strength of Czech accent and the duration of vowels before obstruents**

**Vztah mezi sílou českého přízvuku v angličtině a trváním vokálů před obstruenty**

Praha, 2012 **Vedoucí práce:** Mgr. Radek Skarnitzl, Ph.D.

**Acknowledgments**

I would like to thank my supervisor, Mgr. Radek Skarnitzl, Ph.D., for his time, unfailing patience and kind readiness to give advice on any problem I encountered during writing this thesis.

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 4.6.2012

podpis

## Abstract

This bachelor thesis aims to inspect the presence or absence of a feature called *pre-fortis shortening* in English spoken by Czechs. The term denotes the shortening of a vowel preceding a voiceless obstruent. This feature is known to appear in various languages like Russian, French, Italian; indeed, Matthew Chen even suggests it is language-universal. In English the feature is very prominent and because it affects speech perception, it is even considered a primary indicator of the voicing of the following obstruent. A study included in this thesis examines the extent to which 12 female speakers of Czech English, sorted into 3 categories according to their proficiency in pronunciation, mark the distinction between words like “bet” and “bed” by vowel shortening. The study does not exploit minimal pairs like these, but vowel-obstruent sequences taken from long read passages of BBC news. The the fortis/lenis character of the final obstruent contained in these sequences is a discerning parameter which always separates the data into two groups, the vowel durations of which are then compared. Statistical analysis of the data showed that contrary to the expectations, speakers with native-like pronunciation were not the ones who displayed the most massive usage of pre-fortis shortening. Possible reasons for this unexpected behaviour are suggested in the thesis.



**Key words:** Czech English, vowel duration, obstruent, voicing, pre-fortis shortening

# Abstrakt

Předkládaná bakalářská práce se zabývá tím, zda se v angličtině českých mluvčích vyskytuje zkracování vokálu před neznělým obstruentem. Je známo, že u různých jazyků, jako je ruština, francouzština a italština, je takovéto zkracování běžné; Matthew Chen dokonce uvádí, že je tento jev jazykově univerzální. V angličtině je natolik výrazný, že má vliv na percepci řeči a je považovaný za primární ukazatel znělosti následujícího obstruentu. Studie obsažená v této práci zjišťuje, nakolik 12 českých mluvčích angličtiny, rozdělených do 3 kategorií podle úrovně výslovnosti, dosahuje odlišení anglického "bet" od "bed" zkracováním prostředního vokálu. Tato studie nezkoumá minimální páry jako bet/bed, ale spojení vokál-obstruent v plynulé řeči, k čemuž slouží nahrávky čtených zpráv BBC. Získaná data jsou rozdělená do dvou skupin podle fonologické znělostní povahy obstruentu, který následuje za vokálem. Trvání vokálů se potom porovnávají podle příslušnosti k "fortisové" či "lenisové" skupině dat a také na základě toho, do jaké kategorie výslovnosti patří daný mluvčí. Statistická analýza dat ukazuje, že mluvčí s téměř bezchybnou výslovností nepoužívají zkracování vokálu před neznělým obstruentem v nejvyšší míře, jak se původně očekávalo. Závěrečná část této práce navrhuje možné důvody pro takovéto zjištění.



**Klíčová slova:** česká angličtina, trvání vokálů, obstruent, znělost, zkracování vokálu před neznělým obstruentem

# Contents

# 1 Introduction

The present thesis is focused on how accurate Czech speakers are in reproducing English as their second language. Every language has features by which its native speakers judge, often on the subconscious level, to what extent a foreign speaker of their mother tongue sounds native-like. There is the issue of vocabulary: a command of suitable equivalents to the words the foreign speaker knows from his language. There is grammar, which can be narrowed down to syntax and morphology, each discipline betraying the foreign speaker's skill to express relations in reality despite, for example, a different word-order or an abundance of auxiliary units as in the case of English, an analytical language. Last but not least, there is the phonological and phonetic level responsible for the unmistakable and unique sound of a given language.

As a phonetic study interested in measurable empirical data, this thesis is concerned with the temporal characteristics of speech. Temporal patterning can be identified within all the above-mentioned linguistic disciplines, as will be evident from the theoretical part titled "Temporal organization of speech". The first section of this part (Chapter 2) deals with individual steps of sentence production and how the mental map of temporal patterning is modified within each of these steps. The second section focuses more closely on the phonological component of speech, explaining temporal characteristics of individual classes of phonemes. The third section explores the effects of phoneme combination, namely pre-fortis shortening, a phenomenon found in vowels preceding voiceless obstruents, having a distinguishing function. Chapter 2 is concluded by a remark on perception in relation to pre-fortis shortening.

Chapter 3 presents the empirical part of this thesis, consisting of a study carried out in order to find how Czech speakers of English employ pre-fortis shortening to mark differences in the voicing of the following obstruent. The first section of this chapter clarifies the selection of speakers for the study along with the method of measurement and analysis, stating how the recordings of 12 female speakers were processed while differentiating between 3 levels of proficiency in pronunciation. After that comes a section presenting the results and commentary. The last sections discuss the relevance of the presented results and draw a conclusion to the thesis.

# 2 Theoretical part: Temporal organization of speech

In the act of speech, the speaker conveys information to the listener by generating sound sequences that comply with a set of rules recognized and accepted by both participants. As speech develops in time, one area of these rules concerns segmental duration. The duration of a speech sound is the time that elapses between its onset and its offset. These two moments can be detected as they physically happen, they can be viewed in oscillographic and spectrographic images of speech, and the duration between them measured with great accuracy (automatic segmentation is being constantly perfected and its occasional imprecision can be corrected manually by a skilled phonetician, see e.g. Machač and Skarnitzl, 2009: 12). The human ear and brain which cooperate in the process of perception are not as perfect as a computer. However, it is true that varying the duration of a speech sound in one position may result in the perception of a different piece of linguistic information altogether. A slight but *noticeable* change in the duration of a sound may be an important perceptual cue for the listener, signalling the lexical or syntactic identity of its higher unit, the word.

Interest in these perceptual cues has been listed by Crystal and House as one of the motives for studying speech signal timing. Other motives are e.g. the necessity for "developing durational rules for synthesizing intelligible and natural-sounding speech", or the aim to solve "the problems of automatic analysis of speech for speech recognition, speech understanding, word finding, etc." (Crystal and House, 1988a: 1553-4). These are goals that still shape the character of present-day phonetic research, because it is not an easy task to describe in complexity which temporal aspect of running speech has greatest, medium and least influence on what a person perceives as natural. That is due to the fact that we repeatedly have to confront the objective with the subjective: statistical studies of speech sound durations yield results in exact numbers. On the basis of these, hypotheses can be formed, but have to be proved or disproved by experiment, which involves people participating in recording speech material and in listening tests. This is where the subjective comes in: people are not machines and their performance can be affected just by the circumstances of being watched and recorded. However, when a large number of speech samples are taken, general tendencies can be observed. This inspires new hypotheses and brings speech researchers closer to the core of the problem in focus. It seems to be a concentric movement with one universally applicable rule to be discovered at the centre, only in case of the temporal patterning of speech the "cores" are numerous and combined.

## 2.1 Factors relevant for sentence building

As Nooteboom (1999) says in the part of his study concerning rhythm, there are so many different factors simultaneously influencing the perceptible temporal patterns of speech that it makes life hard for speech researchers who want to give a systematic account of them (Nooteboom, 1999: 15). If we leave consonants aside for now and discuss vowels only, we find that several factors have been shown to have direct influence on their durational realizations.

Van Santen (1992; in Nooteboom, 1999: 17) listed seven factors having quantitative effects on vowel durations:

- Vowel identity
- Identities of the surrounding segments
- Position of the vowel in the syllable
- Position of the syllable within the word
- Stress status of the syllable
- Position of the word in the sentence: the effects of phrase boundaries
- Accent status in the word

We shall discuss some of them below. They correspond nicely with what Klatt (1976) develops in his study on segmental duration in English. He names several phases the speaker goes through in building a sentence, showing that speech timing can be specified within each of these consecutive steps. The seven objective factors named above can be found as dominant in all the respective steps. They affect, more or less directly, the physical realization of speech, which then has a perceptual impact on the listener.

First, the speaker maps out a mental plan of what he is going to say: he makes semantic, lexical and syntactic decisions. Then there comes the phonological component (which means applying the knowledge of which speech sounds may or may not follow which); phonetic rules concerning the potential and mobility of the articulators are implemented next, and the production of speech is finished by articulatory and acoustic transformation into actual sound (Klatt, 1976: 1209). A similar model, proposed by Francis Nolan in 1983, is illustrated in Figure 2.1.

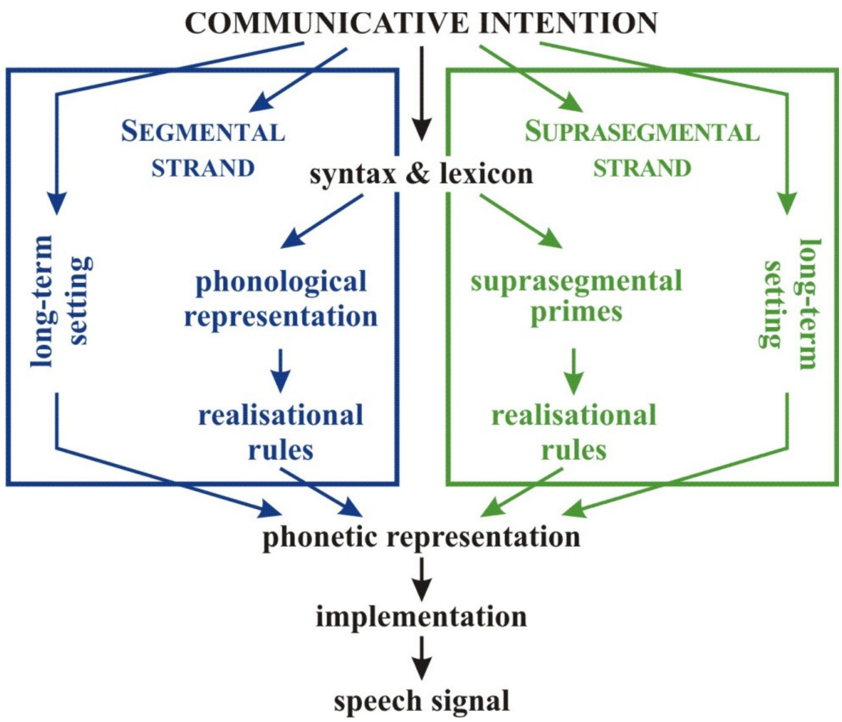


**Figure 2.1.** The process of speech generation, adapted from Nolan, 1983: 30*ff.*

After looking into the relationship of physical segment duration and perceived rhythm, I shall briefly inspect these building blocks, eventually centering my attention around the phonological and phonetic component.

Even before the "shaping force" arising from the language itself takes part in sentence building, there are extralinguistic factors that can predetermine speech rate. Much depends on the situation in which communication happens: whether it is a leisurely conversation or a time-stressed phone-call, whether the communication partner is a child, an adult or perhaps a wide audience, etc.  Furthermore, different emotional states correspond with different speaking rates; Klatt says that anger, fear and sorrow are generally reflected in slower speech (Klatt, 1976: 1210). Of course, we could argue that there are situations in which anger or fear, on the contrary, make speech quicker, because they are excited, not inhibited states of mind. The acoustic correlates of different affective states are being explored by present-day prosody (see e.g. Banse and Scherer, 1996).

One of the main domains explored by prosody is rhythm. It resists firm scientific grasp, because its evaluation is strongly subjective. Listeners are usually able to tap their finger in time with the rhythm of an ordinary sentence. They place the taps approximately synchronously with the onsets of stressed syllables (Rapp, 1971 and Allen, 1972; in Klatt, 1976). It may seem as a positive clue to how the perception of rhythm works, but the taps do not occur in all predicted places – with all stressed syllables. That can be partly due to the fact that there are different types of syllabic stress, which will be discussed below. It would help the creation of universal models of speech timing if the distribution of taps (indicating perceived rhythm) were systematic, but it only suggests that rhythm is an important organizing component of speech, in that listeners are sensitive to it, that it helps them process what is being communicated (Nooteboom, 1999: 10).  Volín (2010) similarly states that the perception of speech is “relatively effortless as long as the temporal structure of the incoming units is predictable” (294). Predictability in speech is very important, because it saves the brain much energy. As it is commonly known, the brain’s activity can be measured in quasi-periodical waves, and it is highly favourable when the information input likewise comes regularly; that way it excites neural resonance more easily (Grosberg, 2003; in Volín, 2010: 294). The listener accustoms the pulses of his attention to the rhythmic pattern of the speaker, being most active when an essential piece of information – e.g. an autosemantic word bearing lexical meaning – is likely to appear. Nevertheless, despite these fairly logical assumptions, we must admit that the rhythm of speech is based on perceptual illusions whose nature is not yet entirely understood (Volín, 2010). It seems that just as a cooking recipe is refined and perfected over years to guarantee the best gastronomic result possible, rhythmical speech is a habit people have improved and developed over time for communication to be most effective.

Although it has been said that rhythm is a perceptual illusion and as such is rather a "black box", it is understandable that there have been attempts to create sophisticated models of rhythm (Lee and Todd, 2004, in Volín, 2010). When a researcher wants to base such a model on empirical data, he has to test a phenomenon which is at the same time perceptually relevant and measurable. Huggins (1972) was interested in just noticeable differences (JNDs) in segment duration. He showed in his study that when subjects based their judgements on changes in perceived stress and rhythm, they were able to detect smaller changes in duration (Huggins, 1972: 1270). That is to say, when they paid attention to whether the sample sentence agreed or did not agree with how they would possibly rhythmically produce it themselves, they were readier to mark an effect of a durational change as unnatural. Huggins also found out that when he simultaneously increased the duration of one segment and decreased the duration of an adjacent segment, the subjects were less sensitive to these two alternations than they were only to one of them, because these alternations were complementary and did not disturb the temporal intervals between stressed syllables, i.e. the overall perceived rhythm (Huggins, 1972b, in Klatt, 1976: 1218).

In his investigation of just noticeable differences, Huggins reached a conclusion that for a listener to perceive a difference in duration, the segment - the author analysed /p ʃ m l ɔ/ - has to be at least 20 ms longer or shorter (Huggins, 1972a; in Klatt, 1976: 1218). Klatt proposes a JND of 25 ms, arguing that Huggins based his experiment on repeating test sentences to the listeners a large number of times, so that the listeners developed a reference pattern against which they judged the changes in duration. Once they familiarized themselves with the contents and form of the sentence and there were no more unexpected stimuli in these respects, they were able to pay much more attention to slight changes in rhythm. However, that does not simulate normal conditions in human communication, therefore a new experiment was carried out by Klatt and Cooper (1975; in Klatt, 1976), in which the authors altered the conditions by introducing a smaller and better mixed-up set of experimental trials, and concluded that only changes in duration exceeding 25 ms are relevant for perception.

After what has been said, it is not surprising to state that two utterances having physically different patterns can be perceived as having the same pattern. It is worth mentioning that English is said to be a stress-timed language, which means that prominent syllables, those carrying primary stress, occur at roughly regular intervals. Klatt (1976) doubts the notion that speakers organize unstressed syllables with the purpose of fitting them into equally measured intervals, only to serve a certain aesthetics of language. He argues that this phenomenon is given by English grammatical rules for stress-placing. However, Nooteboom (1999) managed to demonstrate a related phenomenon on reiterant speech made up of Dutch-spoken nonsense words, which eliminates at least the role of lexically-determined stress. These words were formed by identical syllables, ranging from 1 to a sequence of 4, having stress on the first. What Nooteboom demonstrates here is called *compensational shortening*: the anticipation of many syllables leads to shortening the most prominent one which should be longest. If we accept this proof given by nonsense syllables and extend it upon all syllables within a rhythmic unit, it can be said that a speaker has a universal tendency to modify the rhythmic pattern of what he is about to say so that a unit, marked at the beginning by a stressed syllable, does not change its duration chaotically with every other additional syllable. Regular rhythm can be considered an aesthetic requirement, after all. It is evident from other areas of human activity as well: music, dance, etc. This inclination towards regularity is closely connected with the fact that each of us listened to a mother's heartbeat for a total of nine months.

Let us now turn to the semantic, lexical and syntactic decisions the speaker makes when building a sentence. The first two are manifested by placing stress on syllables. This has perceptual effects, some of which we have described above. It can be agreed that syllables are the basic units of timing in speech (Nooteboom, 1999: 13). The influence of syllabic stress as such on segmental duration has been discussed by Crystal and House (1988b). Their analysis was carried out on vowels in stressed, semi-stressed and unstressed syllables taken from the readings of 6 speakers. It showed that the difference between the duration of an English vowel in a stressed and unstressed syllable is 70 ms on average, which is quite substantial, at any rate 45 ms above the conventionally accepted JND.

First, let us consider durational changes which the speaker makes in order to draw attention to words conveying new or most important information. These changes result in the so-called contrastive stress. Emphasis is put on a certain word by increasing the duration of the whole unit by up to 20 %. In specific examples given by Klatt, this effect of emphasis (or contrastive stress) is also achieved by varying the fundamental frequency (*ibid:* 1210), i.e. by giving intonation cues. Huggins (1972) also observed that varying segment duration can cause shifts in perceived stress, even though the stress is not on the syllable that is being changed. In his experiment, it led to deciphering a different meaning altogether. He provided a sentence "The hostel for paupers has two moody managers", in which he varied the duration of /m/ in "moody". When the /m/ was longer than the original, the listener heard **two (moody managers)**, with stress on **two**. When the /m/ was shorter, the phrase became **(too-moody) managers**, with stress on **managers** (*ibid:* 1276). This is another considerable perceptual impact of segment-duration variation. Not only did stress undergo a shift, but also an imaginary boundary called the *juncture*, which splits the word stream into prosodic units of words that belong together and create a semantic whole. When these units are reorganized, they automatically inspire a different association of meaning in the listener.

Lexical decisions are connected to durational changes in a similar way as semantic decisions are. Lexical stress, however, differs from contrastive semantic stress, because its distribution is less up to the speaker and more tied with linguistic rules. It is governed by the speaker's communicative intention only as far as word choice is concerned; from there it has to obey the prescribed stress pattern, so that the word signals the correct lexical meaning or morphological category. This is the way that the difference between /kənˈtrækt/ (verb – e.g. contract an illness) and /kɒntrækt/ (noun – an agreement), /əˈdʌlt/ (adjective) and /ˈædʌlt/ (noun) is successfully marked. The knowledge of prescribed lexical stress placement is something that marks a good speaker of English (or generally, a language; although for example Czech is less tricky in this respect, for stress is generally placed on the first syllable of a word). When a non-native speaker of English stresses the incorrect syllables in words, it lowers their comprehensibility and may even have undesirable sociolinguistic effects on the listeners, in the sense that they may subconsciously classify him as less competent. What is more, obviously, it disrupts the rhythmical flow of his speech.

To refer back to rhythm once more, lexical stress can sometimes be subject to the above-discussed tendency towards *eurythmy* – a "pleasant" alternation of prominence within a series of syllables. Let us consider the word "thirteen". If pronounced in isolation, primary lexical stress falls on the second syllable: /ˌθɜːˈ**tiːn**/ The first syllable bears secondary stress (it is not completely unstressed, otherwise the vowel would be a schwa). If the word "thirteen" is part of the phrase "thirteen men", the canonical way, /ˌθɜːˈ**tiːn**ˈ**men**/, would lead to two consecutive stressed syllables, or stress clash. That is why the lexical stress tends to shift backwards to make space for a lesser stress between two major stresses: /ˈ**θɜː**ˌtiːnˈ**men**/ (Giegerich, 1992: 277-8). The sequence achieved is then strong-weaker-strongest, which is also easier to pronounce.

In turning our attention to the syntactic factors bearing on sentence building, we are gradually moving from the area of subjective to the area of objective language features. Within this range of factors, several authors looked into what a syntactic pause does to segmental duration. A syntactic pause marks the boundary of a constituent linguistic unit: a phrase, a clause, a sentence. It is most often a true physical pause; called for by the bare necessity to breathe. Speakers tend to make these pauses in logical places, so that the flow of coherent information units (syntactic phrases) is not broken up unsystematically. This is not an absolute rule, of course, it is not difficult to find pauses in natural speech whose placement is, from a strictly syntactic point of view, illogical or unjustified.

Speech sound contexts around these pauses, especially prepausal contexts, have been explored by Crystal and House (1988b: 1577) and by Klatt. They were interested in a phenomenon called *prepausal lengthening*. Pragmatically speaking, this term could be argued to be inaccurate, because it creates the impression that a speaker takes care to lengthen a syllable before a pause, when in fact it is how overall exhaustion of breath and energy demonstrates itself, without any activity on the part of a speaker. Volín and Skarnitzl (2007) have thus suggested a better-fitting term *final deceleration.*

Either way, it has been demonstrated that a syllable (or its nucleus, the vowel) in a word-final position is longer when followed by a syntactic pause than when not. Various authors like Lindblom and Rapp (1973) or Quené (1989; all these authors in Nooteboom, 1999) have also been quoted for noticing this phenomenon and for verifying its validity by experiments. Lindblom and Rapp (1973) found that prepausal lengthening also seems to occur in places where a pause would normally be but is not present in the acoustic signal, which is further confirmation of the phenomenon's significance. It was discovered to be strong also because it manifested itself even independently of the presence of stress. The increase in duration found by Crystal and House was 40 %. Klatt's findings in the same contexts were similar, a 30% increase. Lindblom and Rapp noted (in Nooteboom, 1999: 11) that the occurrence of this feature has a cyclic quality: not only can syllabic lengthening be found at the ends of phrases, but also in the final positions of individual words.

## 2.2 Phonological component

We have been treating gradually smaller units of speech. The basic linguistic unit - the word - has been seen as consisting of syllables, which are further divisible into segments – phonemes. This is the level for the discussion of the phonological component of speech, and the dependence of segment duration on the environment of other segments.

Each segment has its inherent duration. Klatt (1976: 1213) pointed out that the vowels /ɪ ɛ ʌ ʊ/ are generally shorter than other English vowels. Crystal and House (1988a) specify these other vowels as /i e æ ɑ o ɔ u/ and call them long (tense), as opposed to the short ones (lax).[1] It is worth noting that the vowel system presented in their study is American; in British English, /e/ would be realized as the diphtong /eɪ/. O'Shaughnessy (1981, in Crystal and House, 1988a: 1561) noted a tendency for vocalic duration to vary inversely with vowel height. Crystal and House put this hypothesis to test on long vowels and found out that high long vowels /i u/ are truly shorter than mid and low long vowels (/e o ɔ/ and /æ ɑ/). However, the vowels in the mid section of the vowel quadrangle do not hold the same relation to low vowels, thus it was concluded that O'Shaughnessy's hypothesis can only be applied to the vocalic extremes high/low. This finding has a natural explanation: low vowels are intrinsically longer because it takes more time for the articulators (tongue, jaw) to move downwards into the final position for /æ ɑ/ and back up again.

[1] Compare to the set of Czech short vowels: /ɪ ɛ a ɔ u/.

Before we proceed to discussing consonants, let us clarify some terms and means of dividing consonants we will be using. The first necessary step is to draw a distinction between the **manner** and **place** of articulation, and after that add the aspect of **voicing.**

| | bilabial | labiodental | alveolar | postalveolar | palatal | velar | glottal |
|---|---|---|---|---|---|---|---|
| plosives | p b | | t d | | c ɟ | k g | |
| fricatives | | f v | s z | ʃ ʒ | | x ɣ | h |
| affricates | | | t͡s d͡z | t͡ʃ d͡ʒ | | | |

**Table 2.1.** An overview of obstruents used both in English and in Czech. The palatal plosives /c ɟ / and velar fricatives /x ɣ/, as well as alveolar affricates / t͡s d͡z/ are marked by grey colour as Czech, not occurring in English.

a) The **manner** of articulation: Table 2.1 above applies to *obstruents,* the type of consonants we will be dealing with primarily. Obstruents are a category containing the *plosives*, *fricatives* and *affricates.* What these have in common is that the speaker forms an obstruction in his/her oral cavity (hence "obstruents") to modify the air flow coming from the lungs, but the type of obstruction varies: it can be a full closure, which is then released with an plosion-like sound (plosives); it can be merely a constriction, bringing the articulators so close that a hissing, turbulent noise is heard (fricatives), or it can be a compromise between the two, starting in a plosive-like manner with a closure, released gradually, maintaining the critical constriction (affricates).

Consonantal articulation can be thought of consisting of three phases, as depicted in Figure 2.2 below.

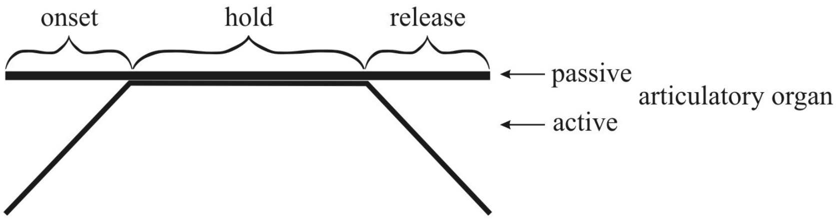


**Figure 2.2.** The individual phases of consonant production: the *hold* stage can either be the maintenance of a closure, or a constriction. In this case, it is a closure, the articulators being in immediate contact.

Depending on the point of view, we call the category of obstruents containing /p t k/ and /b d g/ *stops* or *plosives*. The former label takes into account the articulatory nature of the hold (a complete stop of the air flow), while the latter describes the acoustic circumstance of the release stage, a minor plosion. We shall carry on with the term *plosives.*

b) The **place** of articulation: As the figure above suggests, the speaker makes use of passive and active articulators. The active articulators are the lower jaw and the flexible tongue, which can make contact with several places within the oral cavity. Taken from the front, the first two articulatory places do not involve the tongue: they are the lips (→ *bilabial* obstruents) and the lower lip in contact with upper teeth (*labiodentals*). The *alveolar, postalveolar, palatal* and *velar* obstruents are all defined by the closure or constriction of the tongue against the alveolar ridge, the hard palate and the soft palate (velum), respectively. It is obvious from Table 2.1 that the alveolar ridge is a frequent place of articulation; the affricates are also located there.

There is an additional feature going through the whole system of obstruent place-manner combinations (bilabial plosive, labiodental fricative...), and that is the feature of **voicing**. What is decisive in this respect is the presence of the acoustic signal coming from the vocal folds. It is well visible and detectable in spectrograms as a dark horizontal band of fundamental frequency common to voiced consonants as well as vowels. The voiceless/voiced distinction makes it possible that there is a cognate pair pertaining to each category of obstruent: /p b/, /t d/, /k g/. The following spectrogram and waveform (Figure 2.3) show the alveolar plosives /t d/ in the word "editing", with the absence and presence of phonation apparent in the respective segments.

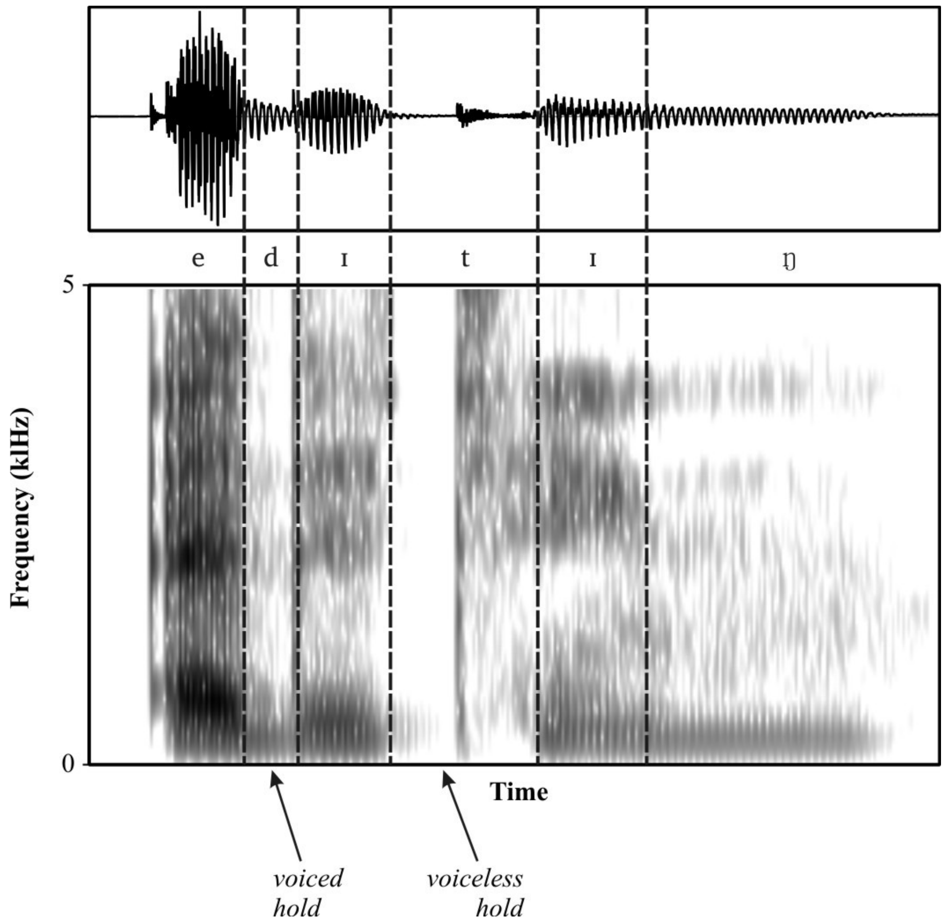


**Figure 2.3.** A spectrogram and waveform demonstrating the presence and absence of fundamental frequency in the alveolar plosives /t d/.

The only category of obstruents that does not contain a cognate pair is the glottal fricative (there is only the voiceless /h/ in English and a voiced /h/ in Czech, which should properly be transcribed as /ɦ/); these, however, will not be in the focus of our attention. Apart from calling the respective obstruents "voiceless" and "voiced", a different concept was developed in the 1960s that took into consideration the energy necessary to produce the obstruent. Voiceless obstruents should then be synonymous with *tense* or *fortis* (as greater force is presumably necessary) and voiced obstruents with *lax* or *lenis* (e.g. Jakobson and Halle, 1962). The disputability of this other concept based on articulatory force will be discussed later.

To draw back on inherent phonological durations, consonants have theirs as well as vowels. Klatt (1976: 1213) argues that voiceless fricatives are about 40 ms longer than voiced fricatives. The reason for this lies in two antagonistic requirements on the vocal tract when a voiced fricative is to be produced: On the one hand, there has to be higher air pressure below the vocal folds than above them in order for phonation to start, which in other words means lower supraglottal pressure. On the other hand, friction (the typical *s*-noise) is possible only with relatively high oral – supraglottal – pressure. Such a demanding feat of maintaining two pressure points at a time results in the fact that voiced fricatives tend to be shorter than their voiceless counterparts (Ohala, 1983; in Skarnitzl, 2011: 59).

There also exists variation in duration within a type of obstruent, depending on the place of articulation. According to Klatt, bilabial plosives are typically slightly longer than alveolar or velar plosives. Crystal and House, however, do not confirm this statement by their data. Quite on the contrary, they show that the total duration of velar stops is longer than the ones of alveolar and labial stops (1988a: 1558). The key to understanding this disagreement lies in separate measuring of the three phases in consonant production.

Crystal and House point out that the average durations for the hold phases of bilabials, alveolars and velars are not very different (although alveolars do display a slight tendency to be shorter than the other two). The release phases, however, are the ones that affect the distinct total durations of the three places of articulation in question. Their average durations increase as the point of contact moves from the lips to the velum. As a result, the total durations amount to about the same for bilabials and alveolars (80 ms), but to substantially more for velars (100 ms). By these findings, Crystal and House disproved the assertions of several other authors who had preceded them in examining the durational tendencies within the category of plosives.

Machač (2006) and Homolková (2009) inspected the temporal behaviour of Czech obstruents within the two separate categories of plosives and fricatives. They confirmed the results of studies such as those cited above: Machač observed that within a cognate pair of plosives, the voiceless segment is always longer than the voiced. Homolková supported this by her probe in the Czech fricatives. Furthermore, Machač confirmed that the alveolar plosives /t d/ tend to be consistently shorter than bilabials, palatals or velars; however, Homolková's alveolar fricatives /s z/ do not show any such visible tendency, as can be seen in Figure 2.4 below:

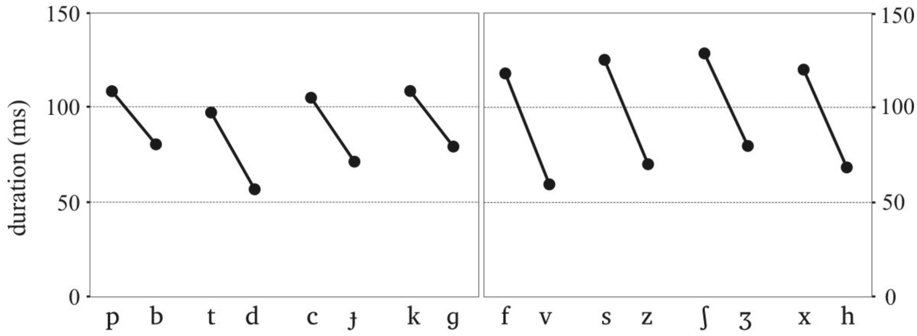


**Figure 2.4.** The mean durations of Czech cognate pairs within the categories of plosives and fricatives, Machač, 2006 and Homolková, 2009 (adapted according to Skarnitzl, 2011: 104).

It is interesting to note that details like these, detectable on the scale of mere milliseconds, are also responsible for the "different sound" of segments like /p/, /t/, /k/ pronounced by speakers of different languages, in our case English and Czech. They are transcribed by the same IPA grapheme, but the English /t/ and the Czech /t/ are simply not the same. When a Czech speaker wishes to sound "native", his ear has to be well tuned and his brain capable of reproducing minute details – of both articulatory and temporal nature – like these and others that will be described later.

Not only do plosives vary in duration depending on whether they are bilabials, alveolars or velars; as has already been hinted at in the term-explanatory section, they are also differentiated by the articulatory feature of voicing. This feature has been thus suggested as another factor significant for consonantal duration. The hold phase of a voiceless plosive has been proved to be much longer than that of a voiced plosive (see e.g. Nooteboom, 1999: 14). Leigh Lisker analysed cognate pairs like "rapid-rabid" or "stable-staple" and found that the hold phase of /p/ was 120 ms, while of /b/ it was 75 ms. It is noteworthy that when the duration of the hold phase of /p/ was shortened under 70, listeners evaluated it as /b/, which means that the distinction has a perceptual significance (Lisker, 1957; in Skarnitzl, 2011: 103). Machač and Skarnitzl (2007) demonstrated on Czech material that /p/ and /k/ are by about 1/3 longer than their voiced counterparts /b, g/; /t/ is on average by as much as 2/3 longer than /d/ (538). Apart from this, Figure 2.5 below also depicts what portion of the total plosive duration is the hold and what is the release.

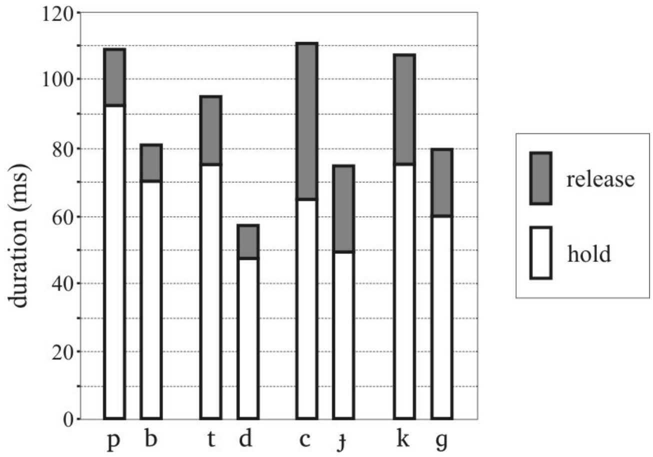


**Figure 2.5.** The durations of Czech plosives with hold and release phases distinguished, Machač, 2006: 36 (adapted according to Skarnitzl, 2011: 105).

## 2.3 Speech sounds in a sequence

Once these inherent segment durations were sufficiently explored by phoneticians, there was but a little step to testing the effect of segment combining. Vowels were discovered to behave in a certain way before certain consonants and that gave rise to the notion of *compensatory shortening*. The description of this effect is based on the presumption that the duration per syllabic unit is relatively constant (Chen, 1970: 146). In reality it means that perhaps the inherent durations of segments pronounced in a sequence are compromised, so that the even flow of syllables is not disrupted. If, for example, the low long vowel /æ/ were adjacent to a preceding velar /g/ - e.g. in the word **gamble**, /ˈgæmbl/ - either the vowel, the consonant, or both should adjust its duration to meet the temporal requirement of the syllable. Another example could include a voiced/voiceless stop in connection with a vowel, which then would be expected to compensate for the different inherent length (short/long). There is also the question of order: whether more compensation will be observed in consonant-vowel or vowel-consonant sequences. Machač and Skarnitzl (2007) tried the hypothesis of compensation either way for the Czech language. They studied PV and VP sequences[2] and arrived at these conclusions:

[2] PV, VP: plosive-vowel, vowel-plosive; "plosive" being a voiced/voiceless stop from the set of cognate pairs /p b t d c ɟ k g/.

- The duration of plosives (related to their voicing status) tends to affect the duration of adjacent vowels. The initiating factor of vocalic compensation is the voicing contrast in plosives.
- The duration of vowels (related to their height) tends to affect the duration of preceding plosives. The initiating factor of consonantal compensation is vowel height. (540)

Both of these conclusions are related to the PV sequence, as compensation tendencies appeared to be stronger in PV than in VP sequences in Czech. The reason for this being the case is that the segments in question are tautosyllabic (they are assumed to belong to one syllable), whereas VP sequences are not (*ibid:* 540). English, however, differs greatly from Czech in this respect – the VP sequence is organized otherwise when the vowel belongs to a stressed syllable: the boundary is placed after the consonant.

### 2.3.1 Syllable boundaries

The search for the best way to locate a syllable boundary has a long history. The general intuitive consensus has always been that a syllable is, at its core, a physiological unit (Krakow, 1999: 23). A number of researchers had tried to find evidence for the presence of syllable boundaries in the movement of articulators, but the features they observed were unsystematic: there was no clear-cut relation between respiratory muscle activity and syllable organization during speech production, nor was there any systematic signal given by the troughs in tongue raising and fronting in isolated two-syllable sequences like /ipi/ (Harris and Bell-Berti, 1984; in Krakow, 1999: 24), let alone in connected speech. Clearly, a different approach to the investigation of the syllable than looking for its physiological basis was necessary. This new approach took advantage of the fact that consonants are associated with different characteristic articulatory patterns (*ibid*: 25), depending on their position in the syllable (whether initial in a CV sequence, or final in a VC sequence). For example, Browman and Goldstein (1995) or Keating (1995a) asserted that the hold phase of the plosives /p t k/ and /d/ is tighter syllable-initially than -finally (in Krakow, 1999: 34). The fact that the initial consonant is well distinguished from the following vowel by the strong closure-aperture contrast brings evidence for the general language preference for CV sequences. It is well worth mentioning that many languages allow *only* CV syllables, or that children's babbling always imitates a CV structure, when they begin producing their first "words" like /ba/, /pa/ and /ma/.

The syllable affiliation of consonants in postvocalic positions (e.g., the affiliation of the second /t/ in "potato" or the affiliation of /n/ in "manner") is, according to Krakow, debatable. In her articulatory studies focused on nasals, she examined the movement of the soft palate in triplets of phrases, such as "see more – seam ore - Seymour" or "hoe me – home E – homey". She wanted to find out whether the pattern of velum movement in the /m/ in the third phrase of both triplets is more similar to that in the first or the second phrase. The /m/ in the first phrase clearly belongs to the second syllable (because it begins a new word), whereas the /m/ in the second phrase belongs to the first syllable; if "potato" or "manner" were syllabified according to the first phrase, they would be transcribed as /pəˈ**teɪ.t**əʊ/ and /**mæ.n**ə/ (the dot symbolizing the syllabic boundary, **CV.C**). If according to the second phrase, they would be /pəˈ**teɪt.**əʊ/ and /**mæn.**ə/ (**CVC.**). Krakow's results unambiguously suggest that the latter is common reality. The presented data show that the middle /m/ in "Seymour" and "homey" are not prominent enough to resemble the initial consonant of another syllable, which is an argument in favour of tautosyllabic CVC sequences.

In accordance with what has been said about syllabification, the rule for placing the syllable boundary that this study is going to follow is the one according to Wells (2008: xxvii): Count a plosive to the syllable "to the left" if this syllable is stressed. It is a rule simple enough to follow in monosyllabic words like "base" /ˈ**beɪs**/, but it is also possible to find a desirable context in polysyllabic words like "supervision" /ˌsuːpəˈ**vɪʒ.**n/ or "discovered" /dɪsˈ**kʌv**əd/.

### 2.3.2 Pre-fortis shortening

In English words that will be subjected to examination in this thesis, we shall be interested in compensatory shortening affecting the vowel durations in CVC sequences. The identity of segments surrounding the vowel in question, namely that of the consonant following the vowel, will be the key factor used for interpreting changes in vowel duration.

Only few assertions have been made about vowel duration *after* a consonant in English, but the ones that have been made are in accordance with the results of Machač and Skarnitzl (2007): that vowels after /b d g/ are longer than those after /p t k/ (see e.g. Fischer-Jørgensen, 1964). However, contrary to Czech, most compensation in English is observed in VC sequences, therefore the consonants following vowels are of greater importance to us.

Studies mentioning the influence of consonantal environment on the syllable nucleus go as far as the 1950s (House and Fairbanks, 1953). In 1960, Peterson and Lehiste directed phonetic interest towards the consonant following the vowel. Crystal and House (1988a: 1559*ff*) spoke about "the vocalic *lengthening-before-voicing effect*" and already labelled it as being "well known in English". Chen (1970) had preceded this by a study that even expanded the notion cross-linguistically. He conducted a survey the outcome of which was that vowel duration as a function of the voicing feature of the following consonant is presumably a language-universal phenomenon (Chen, 1970: 139). He compared vowel duration in words that formed minimal or near-minimal pairs (e.g. "lack-lag"; "kilt-killed" pronounced /kɪl**t**/ and /kɪl**d**/). He compared such pairs in the following four languages: English, French, Russian and Korean. In all these analysed languages, vowels preceding voiced obstruents proved to be longer than those preceding voiceless obstruents. The only difference was how massive the lengthening-before-voicing effect was.

In English, it is not only lengthening we talk about. The notion of shortening before voiceless obstruents should not be left out – indeed, it is a second effect called the *pre-fortis shortening*. Chen does imply that lengthening and shortening are two phenomena occurring in different phonetic situations, as can be seen and clarified in the following example: If the vowel /i:/ in the word "bee" is considered neutral and longest for not being affected by any consonant to close the syllable, then the words "beat" and "bead" form two distinct cases opposed to the neutral "bee": in /bi:d/, the vowel is affected by the lengthening-before-voicing effect, but nevertheless is still shorter than in /bi:/; in /bi:t/, the /i:/ is affected by pre-fortis shortening and is shortest of these three cases; phonetic transcription would be [biˑt] to indicate the shortening. Chen compared the duration of the lengthened vowel to that of the shortened vowel and discovered that in English, the latter only reaches 2/3 of the former (0.61), as is clearly visible from Figure 2.6 below. It is notable that of the four languages in question, English demonstrates the greatest difference between the two mean durations. The other three languages show the same tendency, but English has incorporated it in its phonological system as a signal of the voicing feature of the final obstruent.

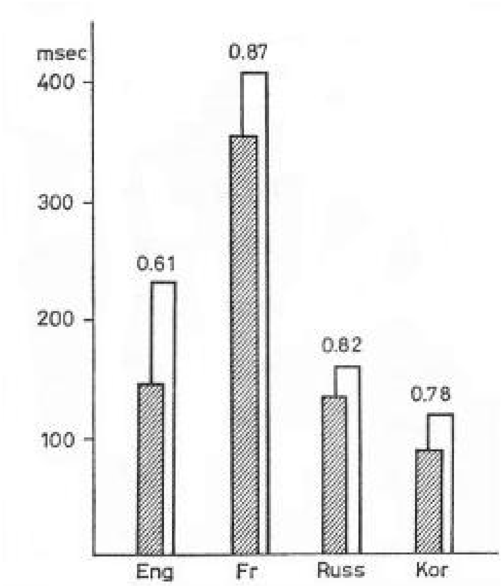


**Figure 2.6.** The results of Chen's comparison of vowel duration in minimal word pairs in English, French, Russian and Korean (see text). Adapted from Chen, 1970: 138.

As the minimal pair /kɪlt/ and /kɪld/ suggests, Chen also examined the behaviour of the vowel when there is an intervening sonorant – the question therefore was whether the lengthening/shortening effect is sustained even across an inserted /m n ŋ l r/ (note that /r/ can only occur in rhotic variants of English). The hypothesis he wished to verify was Halle and Stevens's *theory of laryngeal adjustment* (Halle and Stevens, 1967; in Chen, 1970: 147-8). It proposes that the transition from a vowel to a voiced obstruent requires a delicate change in the vibration of vocal folds, because the obstruent is voiced "non-spontaneously", as opposed to the vowel. This transition takes a little while, therefore the preceding vowel allows for it by extending its duration. Chen expected that, according to the logic of the theory, this adjustment should affect only the single neighbouring segment; however, the lengthening or shortening effect of the consonantal environment was not limited to the immediately preceding sonorant alone, it was observed with the vowel as well (*ibid*: 150).

This is only one of numerous theories Chen introduces in his study and consequently puts to test. His aim is to find a plausible explanation for the empirical fact of the shortening and lengthening phenomena. A very influential theory was voiced by Simon Belasco (1953; in Chen, 1970: 140): Given that the fortis-voiceless obstruent takes more force to produce (as "fortis" suggests), the anticipation of spending more energy on it shortens the preceding vowel. E. A. Meyer (in Lindblom, 1967) took a similar approach: "The temporal organization of speech sounds is determined by the amount of physiological energy that is consumed in producing them" (in Chen, 1970: 141). It is implied here that each syllabic unit takes a constant amount of energy to produce, so that the energy expended to produce the vowel varies inversely to the energy consumed by the following consonant. But electromyographic recordings, monitoring the relevant muscle activity, did not support Belasco's theory of articulatory energy expenditure, as was eventually argued by Zimmerman and Sapon (1958).

Another theory discussed by Chen brings us back to what was the point of departure for Machač and Skarnitzl (2007) in their study of Czech PV sequences: a theory of temporal compensation in order that syllable duration remains more or less constant. It is true that the hold phase of voiceless plosives is longer than that of voiced plosives. However, vowel duration did not register any shortening when Chen measured the duration of words ending with a consonantal cluster. The words "pike" /paɪk/ and "piked" /paɪkt/ should differ in duration, but Chen shows that /paɪkt/ is by no means compressed to be as short as /paɪk/; on the contrary, the consonant cluster stretches the syllable (Chen, 1970: 147). In the strictest sense, the theory of temporal compensation within syllables does not apply to normal human communication. There are several intervening extralinguistic factors, some of which we have already discussed, such as subjective speech tempo and rhythm, making a perfect mathematical speech lay-out impossible.

After this series of disproved hypotheses, Chen comes up with his own theory. According to him, the difference in the duration of vowels preceding consonants depends on the speed of the transition from the vowel to the consonant closure (1970: 152). This transition is faster when the speaker anticipates greater effort in holding the articulators together (e.g. lips with /p/). Such greater effort is necessary in the case of fortis (voiceless) plosives, because pronunciation with an open glottis leads to greater building up of intraoral pressure behind the articulators. Lenis (voiced) plosives do not require such effort, therefore the transition can be slower. Compare the transition phases in Figure 2.7 below.

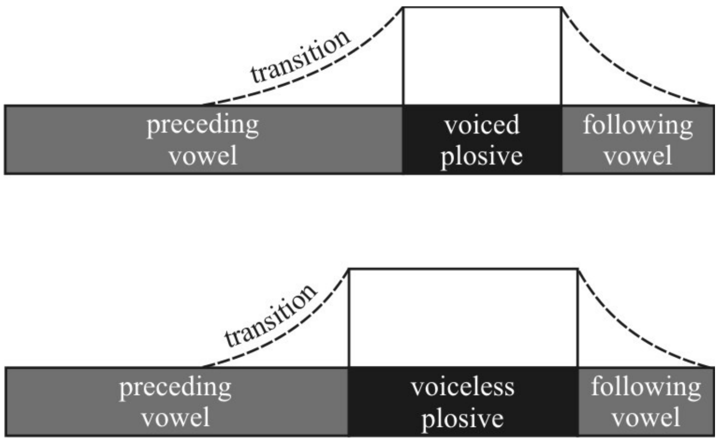


**Figure 2.7.** Two VCV schemes showing the durations of transition phases from a vowel to a voiced and voiceless plosive. From Skarnitzl, 2011: 109; adapted from Chen, 1970: 153.

## 2.4 The perception of temporal patterning

The way we have discussed the temporal organization of speech so far calls for putting it into a larger frame. We may roughly set out speech production as comprising three separate "stages". Let us recapitulate the first two we have been dealing with chiefly:

There are certain factors – rules and restrictions – of speech that are given by language, which are reflected in planning speech in the mind (semantics, lexis, syntax) and then producing it by the articulators. On the border of these two actions, first mental and then physical, we find the area of phonology and phonetics, because individual speech sounds are predetermined both by what is acceptable for the linguistic system and by the articulator mobility. For example, there are consonantal clusters nearly impossible to pronounce (e.g. the final cluster in "sixths"/sɪksθs/), which leads to omission, simplification, sometimes to the point at which the listener is unable to place the word in the language system he has in his mind.

If the first stage is the **mental planning** of speech and the second its **physical realization**, the third is the domain of the listener – **perception**. Communication is successful only when the signal reaches the auditory system of the listener and his/her brain deciphers it the intended way. Within this "stage" we have discussed much about rhythm and its effect on the perception of speech, and have concluded that perception is the most difficult stage to be scientifically defined. Similarly, the existence of variable vowel duration before obstruents has been proved in the first and second stage, but its unquestionable effect on the listener has not been confirmed by any vast number of studies. Suffice it to say, however, that experiments carried out by some phoneticians involving listening tests are coming closer and closer to specifying exactly which portion of the vowel preceding an obstruent (voiced/voiceless stop) has the heaviest influence on the listener's decision "voiced/voiceless" (apart from the voicing itself). Chen has come close to locating the portion as the transition phase from a vowel to a plosive. In a remark on perception, he asserts that speakers have simply developed the different duration ratios of vowel:voicelessP and vowel:voicedP in order to maximize perceptual distance between them. It may be seen as a perceptual device already having a distinctive function in the phonological system of the English language. In the case of word-final post-vocalic voiced fricatives, which are becoming devoiced, vowel length may even become a primary perceptual cue for the voicing contrast (Denes, 1955; in Klatt, 1976).

To conclude the introductory part, we have described factors influencing temporal structure of speech, from speech rate and rhythm to features operating on the segmental level. We also pointed out that not being acquainted with the prosodic rules specific to a given language may have a negative perceptual effect on listeners native to that language. Not only can it lower the speaker's comprehensibility, it may also lead to the listeners' involuntary negative evaluation of the speaker as a person. The objective of the present thesis is to link acquisition of English by Czech learners with one feature related to the temporal structure of speech, *pre-fortis shortening*.

Our research conditions were not connected with any concrete sociolinguistic situation, however, we did work with the fact that there are certain features in a non-native speaker's use of language that induce different judgements on the level of their proficiency.

To be more specific, we decided to look into the degree of employment of pre-fortis shortening by Czech speakers of English in VC sequences, where C is either a fortis or lenis obstruent. The hypothesis we set out with was that speakers whose pronunciation is judged to be near native-like would demonstrate a very clear-cut difference between vowel duration preceding a voiced and a voiceless obstruent. With speakers who did not rate as native-like, the distinction would then be more blurred, certainly not nearing the 0.61 ratio given by Chen (1970).

# 3 Empirical part: A study

## 3.1 Method

In my thesis, I have been working with recordings that had been pre-processed in the sense that the speakers of Czech English had been classified according to their pronunciation. They had been placed in three separate categories, A, B and C (A signifying native- or near native-like pronunciation, C worst – having a strong Czech accent – and B being the widest category of ambiguous cases).

Defining which speakers are nearly proficient in their English pronunciation and which are not required a systematic approach. Skarnitzl *et al.* (2005) carried out a series of listening tests in two consequent stages. The available data comprised 128 recordings of 39 male and 89 female students of English studies. Texts they read were BBC news bulletins of about 500 words organized into 6-7 paragraphs, which the students had had enough time to read and prepare. Using longer, communicatively meaningful texts is in a way innovative in research of temporal features like *pre-fortis shortening*, as experiments usually use target words embedded in carrier phrases, such as "Please say ____ again."

The recordings were obtained in a soundproof booth with a studio electret microphone IMG ECM 2000 and digitized at the sampling rate of 22,050 Hz (*ibid:* 2). During the experiment, 26 of these recordings were excluded due to their bad quality or a non-Czech origin of the speaker. The study ran in two evaluative stages, the first of which included the authors' expert phonetic evaluation. After that, the recordings were presented to 10 native speakers of English and 10 Czech proficient speakers of English. These respondents had been asked to focus only on pronunciation, disregarding slips of the tongue, etc. The result was that evaluation from Stage I coincided very well with evaluation from Stage II, which means that the standards for judging foreign accentedness were roughly similar, although, obviously, there were several inter-respondent differences (on a scale of 1-5, some may have been reluctant to give the grade 1 or 5).

Thus sorted and confirmed, the set of data was ready for further processing, which became part of the research discussed in this thesis. 12 recordings were selected to represent the three categories, A, B and C - 4 female speakers from each. The specific texts they read varied, but that was irrelevant, because in all these texts there was an abundance of the desired VC contexts.

First of all, the material was divided into separate breath groups (BG), with respect to the individual speakers' prosody. This simplified locating phrase-medial and phrase-final words later on. Secondly, these breath groups were processed by the Penn Phonetics Lab Forced Aligner (P2FA, Yuan and Liberman, 2008) that provided the material with rough but relatively precise estimation of segment boundary placement. This had to be manually corrected in Praat (Boersma, P., Weenink, D., 2012) by the author, but only in the relevant CVC contexts. To this end the author made use of principles listed in a handbook by Machač and Skarnitzl (2009). It proved useful especially in cases of segments with a gradual decay of formant structure, as it appears in fricative-vowel or vowel-fricative sequences. The boundaries were placed near the midpoint of the transition area (Machač and Skarnitzl, 2009: 45). As to which boundaries precisely were corrected, they were the ones defining the extent of the vowel in VC sequences, so that the vocalic duration could be measured exactly.

As this introduction has shown, there are many factors influencing the duration of a particular speech sound, one of them being speech rate. In order to eliminate the effect of speech rate on the duration – in other words, to prevent variations in speech rate from blurring the values of vowel duration – raw vowel durations were normalised against local speech rate. For each target word, we took into account one stress group on each side and normalised the duration of the vowel against the number of syllables contained in the 3 stress groups. For example, the underlined portion in the following sentence was used to calculate normalised duration of the vowel /i:/ in the target word "meet": Mister Perez is also exˈ<u>pected to ˈ**meet** the ˈnew Paleˈ</u>stinian Interior Minister.

All the necessary data was then extracted from the material with the help of a script. We retrieved a total of 878 vowels in numerous VC contexts as well as prosodic positions, while specifying the type of the following obstruent as fortis/lenis. The study took into account strictly the underlying fortis/lenis identity, disregarding the sometimes imperfect physical realisation of voicing, especially in categories C and B. We were also interested in VSC sequences, where the S was a sonorant /m n l/ (inspired by Chen's experiment regarding the theory of laryngeal adjustment; see p. 27). The sonorant /ŋ/ was ruled out in our material due to low occurrence and /r/ due to the fact that in non-rhotic English, the segment is not physically present, it only causes lengthening of the syllable (e.g. in "border"). Statistical analyses were carried out in the Statistica program, version 7.

## 3.2 Results

The very first statistical result was predictable and only served as an affirmation of our course. Figure 3.1 shows that vowel duration before lenis obstruents is indeed greater than before fortis obstruents; in our case, with all categories (A, B, C) combined, the average difference makes 9 ms. The ratio of the vowel durations in lenis/fortis contexts amounts to 0.82, which is more than Chen's ratio for English (0.61), however, it curiously coincides with his ratio for Russian (see Figure 2.5).

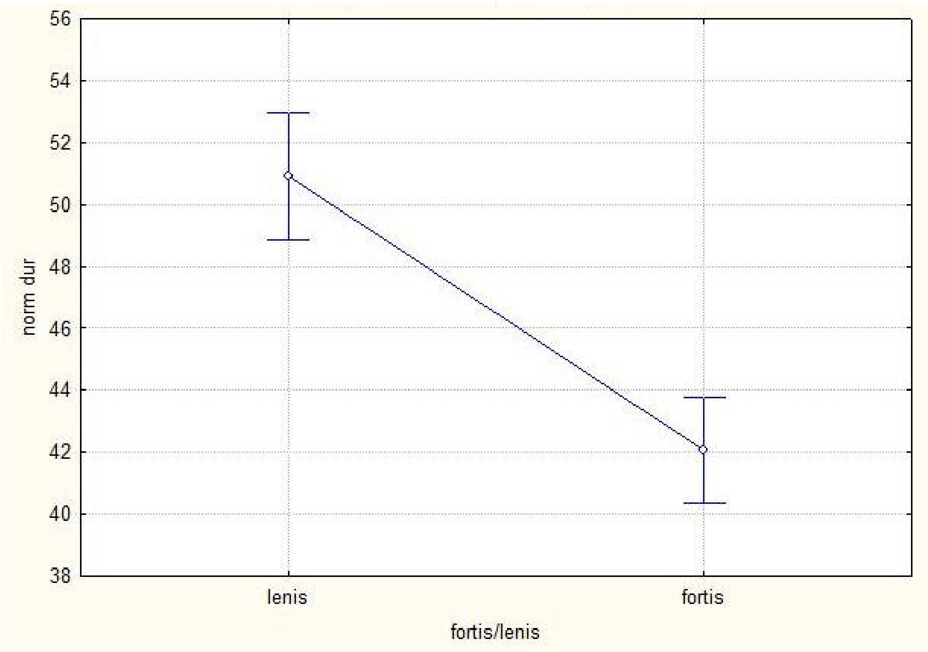


**Figure 3.1.** A comparison of mean values and 95% confidence intervals of vowel durations before lenis and fortis obstruents.

It should be noted here that our data were obtained in two parallel sets: first with the employment of absolute vowel durations, then with the employment of normalised vowel durations. The two sets were compared and as they did not exhibit any striking differences, only the latter was chosen for demonstrating results in the present thesis. This decision was based on the assumption that normalising vowel durations gives a more precise idea of what the relations are between the categories A, B and C; simple auditory experience proved that C speakers speak more slowly and carefully than A speakers. In the rest of this chapter, when we speak about vowel durations, we will therefore always refer to *normalised vowel duration*.

Another thing to mention is the significance of the vertical "I" lines which represent a 95% confidence interval, i.e. a temporal range within which 95 % of all the measured values are located. In other words, the ranges do not represent only the extreme values. These 95% confidence intervals also appear in the following figures and they present the results of ANOVA. It is favourable when the ranges for lenis and for fortis contexts do not overlap, because then the results are also prone to a high statistical significance $p < 0.001$.

After the predictable outcome in Figure 3.1, another factor (A, B, C membership) was added for more detailed analysis. This yielded unexpected findings: the hypothesis that the best English speakers make most use of vowel lengthening or shortening was not confirmed, as is evident from plain sight in Figure 3.2 below:

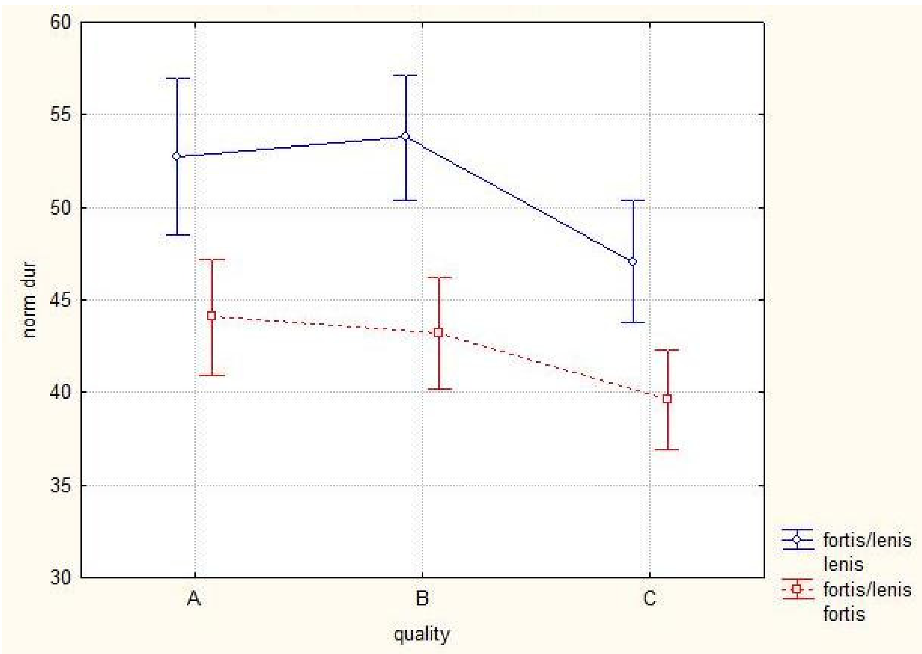


**Figure 3.2.** Mean durations of vowels preceding lenis and fortis obstruents for the individual categories of speakers: A, B and C (referred to as speaker "quality").

If Figure 3.2 were in accordance with our hypothesis, the durational distance between lenis- and fortis-preceding vowels would be greatest in category A and progressively decrease for categories B and C, respectively. As we can see above, our situation is different: while the pre-fortis/pre-lenis duration ratio is – according to our expectations – highest for category C (0.85), it is the speakers in category B which display the lowest ratio (0.81), not speakers in category A (0.83). It is also worth observing that $p > 0.1$ in this case, which seems to render the relationship between vowel durations, speaker qualities and fortis/lenis specifications with no statistical significance. However, interestingly, Tukey's post-hoc tests revealed high significance for the B category results ($p < 0.001$) and significance for the C results ($p < 0.05$).

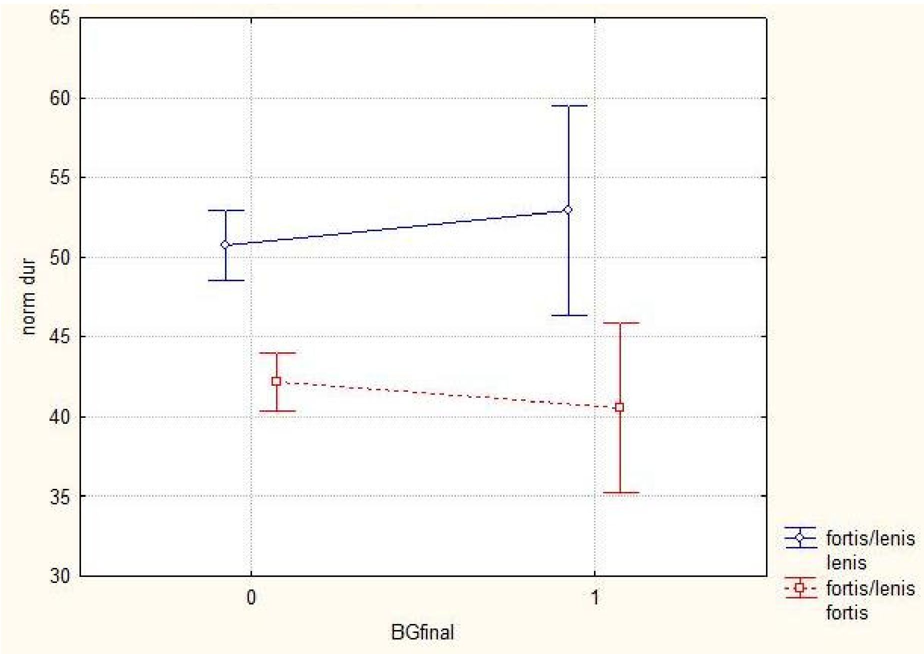


**Figure 3.3.** A chart showing BG-medial (0) and BG-final (1) vowel durations in both contexts (fortis/lenis).

As was mentioned earlier, the recordings were first split into breath groups (BG). At the end of such a unit, speakers are known to slow down and drop their voice, therefore the very last word, when it was used, could be judged as phrase-final. These words formed a distinct group of data (BG "1") and it was expected that the vowel in both contexts (fortis/lenis) would be lengthened. However, it was true only with the fortis-preceding vowels. Furthermore, it is apparent from Figure 3.3 that the 95% confidence intervals are large and nearly overlapping in the BG-final words, which signals great variance of values within the whole group and little difference between the values for lenis and fortis contexts separately (hence $p > 0.1$).

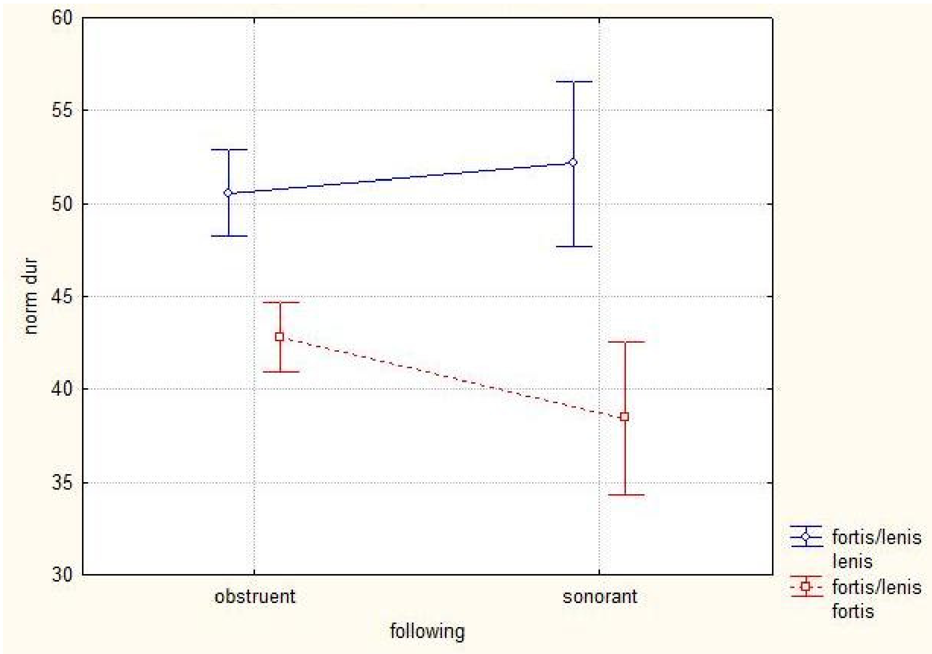


**Figure 3.4.** Average durations of vowels in VC and VSC sequences, depending on the fortis/lenis nature of the final obstruent.

Figure 3.4 above shows the results of ANOVA with the intervening sonorant functioning as an independent variable (comparing cases like "bet" and "bent" as two separate groups of data). At a first glance, it is very similar to Figure 3.3, because the values of vowel durations in VC sequences manifest lower variability than the values in VSC (vowel-sonorant-consonant) sequences, so the first two intervals appear "smaller" than the second two. The figure is more optimistic even, as the ranges for the latter do not overlap: the difference in the duration of the vowel before a fortis and a lenis sound is significant both in the VC and the VSC sequences.

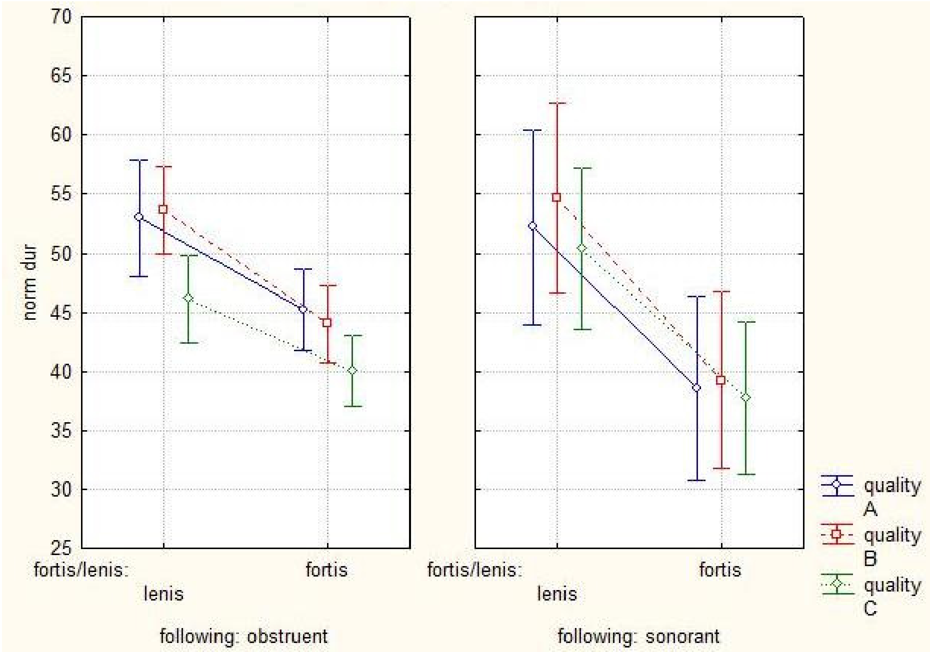


**Figure 3.5.** Mean vowel durations in VC (following: obstruent) and VSC (following: sonorant) sequences, depending on the fortis/lenis context and on speaker quality (A, B and C).

In order to draw upon our pronunciation-quality criterion, it is useful to take a closer look at the distribution of vowel duration values for individual speaker categories (A, B, C). It is apparent from Figure 3.5 that, especially in VSC sequences, the values did not vary much for individual categories. Also, the value ranges overlap very significantly. The only category that seems to make a difference is B in VC sequences (indicated in the figure by the red line: smallest and non-overlapping ranges). Indeed, Tukey's post-hoc tests confirm the significance ($p < 0.05$) of the results for category B, which is the second time the category is prominent this way.

A third case of significance for group B arose when we examined how the "long vowels" (speech sounds corresponding to RP /ɑː iː uː/ and /aɪ aʊ eɪ ɔɪ əʊ/) behaved as opposed to "short vowels" (/ʌ æ e ɪ ɒ ʊ/) in VC sequences. It is well visible from Figure 3.6 that the only ranges that do not overlap are those representing results for long vowel durations of speaker categories A and B. Both these sets of results are significant ($p < 0.05$) and are in accordance with the prediction that the categories containing speakers with better pronunciation would display greater durational differences between vowels in fortis and lenis contexts.

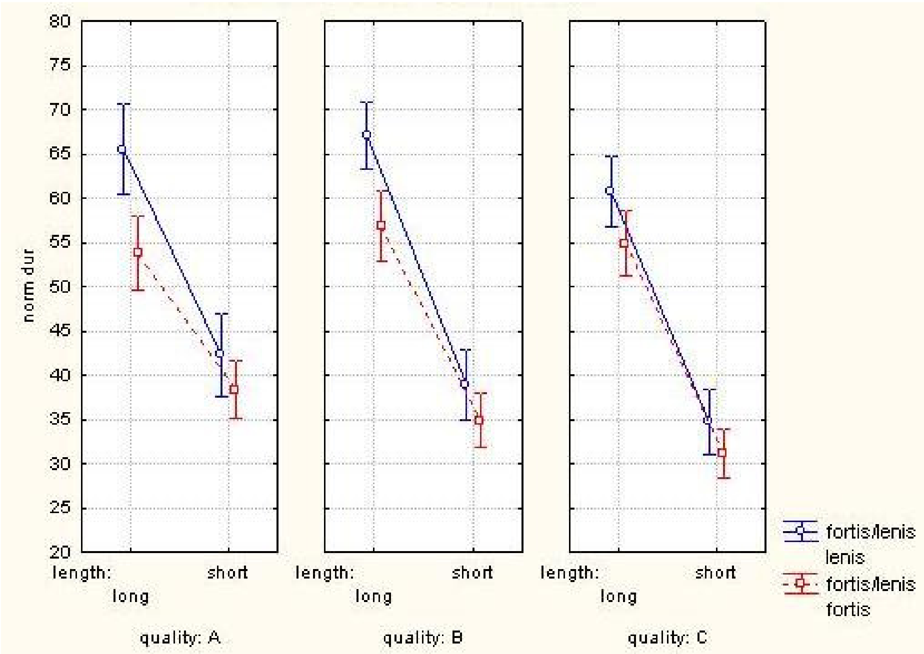


**Figure 3.6.** Mean vowel durations for categories A, B and C depending on the "phonological" length of the vowel (see text).

As far as pronunciation and accent are concerned, we have only discussed the results in the terms of the three categories. We have learned that the analysis of category B yields remarkably significant results. To get a clearer picture of possible internal factors, it is necessary to go one level deeper and examine the individual speakers within the categories.

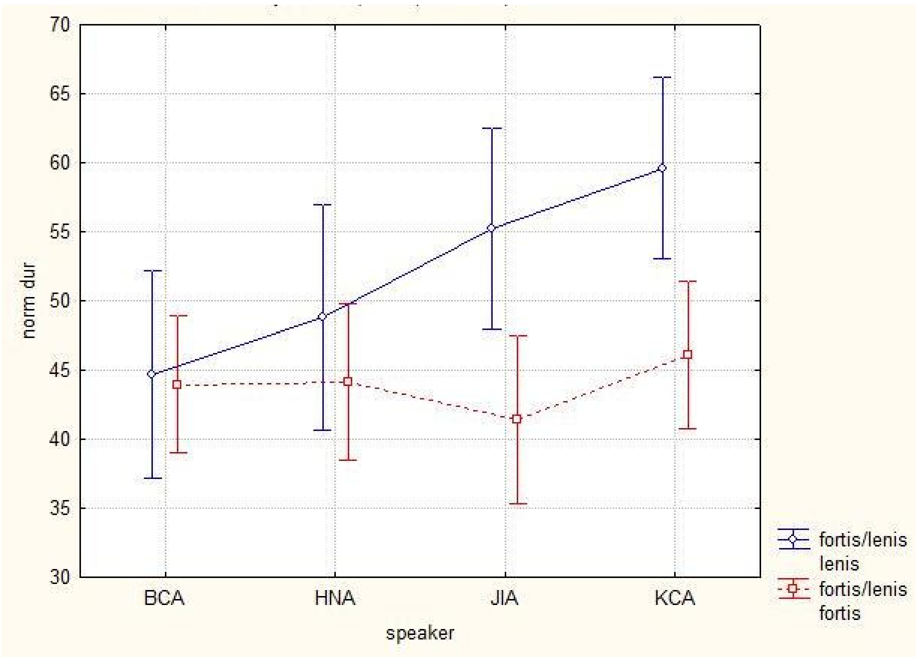


**Figure 3.7.** Mean vowel durations (fortis/lenis contexts) for speakers within category A.

Viewing the speakers within category A in Figure 3.7, we can assume that KCA's vowel durations best comply with the original hypothesis. It is true that auditory analysis confirmed KCA's excellent command of native-like pronunciation (British accent). All "A" values are dispersed into quite wide 95% confidence intervals, however, KCA's intervals do not overlap, therefore her results display a marginal significance ($p < 0.1$), as do the results of JIA, whose fortis-preceding vowels are the shortest of the group. Ideally, the category A would have maintained the largest distance between the values for fortis and lenis contexts.

Speakers within category B were expected to demonstrate a lesser difference between the mean values for fortis and lenis contexts. Figure 3.8 shows that their data was found within tighter ranges around the means than the data in category A. Two speakers (KLSA, KOTA) behaved the expected way, whereas KANA and MILA – especially MILA – showed a wider gap between the values. KANA's results did not prove to be significant in Tukey's post-hoc tests, but MILA's were highly significant ($p < 0.001$). It is therefore obvious that MILA's results caused the significance achieved by the whole category B in the previous results.

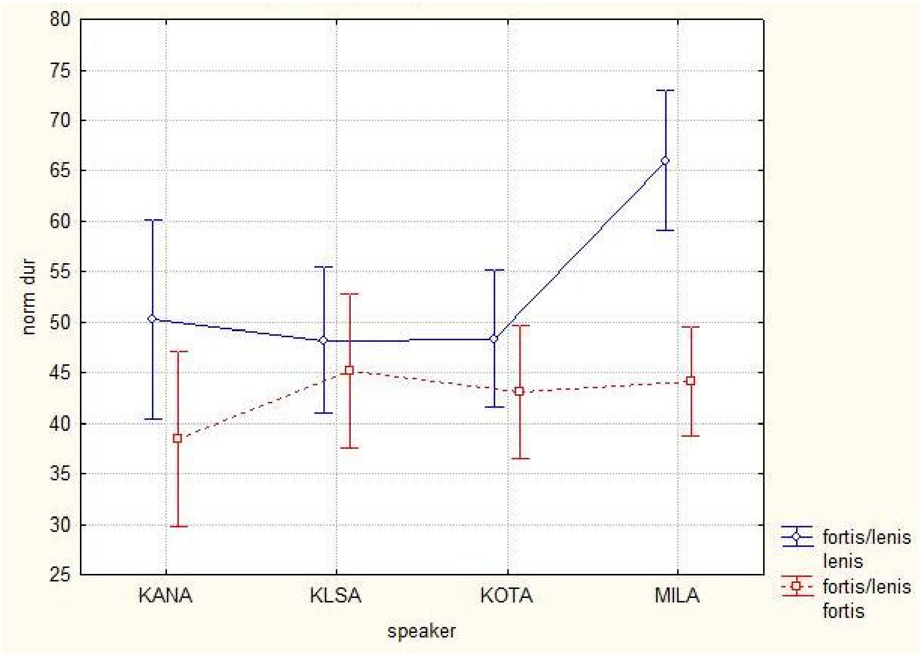


**Figure 3.8.** Mean vowel durations (fortis/lenis contexts) for speakers within the category B.

During the post-hoc auditory analysis, the author discovered that MILA did stand out among the other B speakers – in the sense that she spoke carefully, paying special attention to placing lexical stress and marking it clearly in the temporal domain. The auditory effect was that of prosodic regularity, which could be, disregarding MILA's noticeable Czech accent on the segmental level, deemed almost native-like. However, this was not a feature that could change the overall impression which had been decisive for categorizing the speakers as A, B or C. It could have improved marking the durational difference, as the carefully placed stress coincided with the syllables that were subject to analysis. It may therefore be the case that this speaker's above-average competence in the prosodic domain is reflected in clearer distinguishing in vowel duration before fortis and lenis, such as we would expect from speakers in category A.

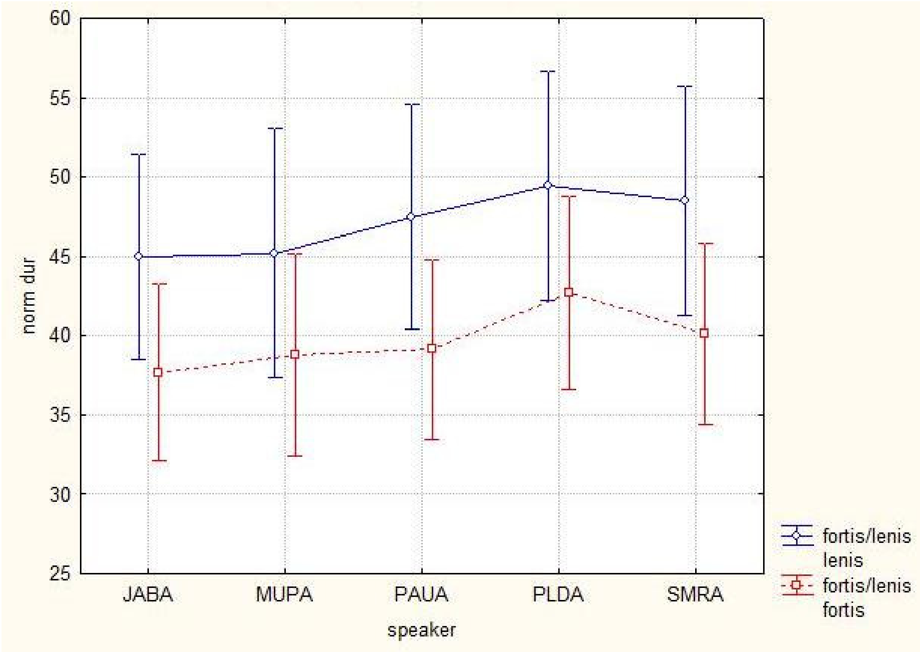


**Figure 3.9.** Mean vowel durations (fortis/lenis contexts) for speakers within the category C.

Category C is interesting in that the data of all speakers are very much like those of KANA from category B – the differences between mean durations for vowels in fortis/lenis contexts are all ca. 10-12 ms (3 ms more than the average difference for all categories combined, see Figure 3.1). That is more than two speakers of the category A achieved. The mean values in Figure 3.9 show that speakers from category C, compared with one another, are most consistent in their pre-fortis/pre-lenis duration ratios, i.e. they all maintain approximately the same spacing between their mean values for fortis and lenis contexts. It is this group where the universal phenomenon, found by Chen in languages like French and Russian, seems to be illustrated most clearly. Most speakers, across all categories, appear to adhere to this universal duration ratio of pre-fortis/pre-lenis vowels (0.8). Three speakers are exceptional – KCA and MILA exaggerate the universal value towards native-English values (0.61 according to Chen), while the duration difference between pre-fortis and pre-lenis vowels in speaker BCA is negligible.

## 3.3 Discussion

The departure point for this thesis was a hypothesis that Czech speakers of English with native-like pronunciation make use of a phenomenon called *pre-fortis shortening* and that those speakers whose pronunciation is marked by a strong Czech accent do not, or very little. We would like to remind the reader that there is what we may call a universal distinction between pre-fortis and pre-lenis vowel duration (with the ratio being around 0.8). English has been shown to emphasize this distinction in order to cue the voicing identity of the final obstruent (with the ratio being around 0.6 according to Chen, 1970). It was therefore expected that speakers of category A would approach native English values, while values in speakers of category C would be in line with the universal tendencies.

The results presented in Chapter 3 digress from our hypothesis. For category C, the ratio was 0.85 and it was maintained very consistently by the individual speakers in the category. Surprisingly, category B yielded a "better" ratio than category A (0.81 and 0.83, respectively), which, however, is caused by the exceptional behaviour of speaker MILA. Her ratio of 0.66 exceeded even the best speakers from category A. The ratio of category A itself was 0.83, with one speaker (KCA) achieving marginal significance, the rest having the groups of fortis and lenis data largely interspersed and approaching the universal values. The differences between the mean values of vowel durations (before fortis and before lenis obstruents separately) lie well below the just-noticeable-difference reported by Huggins and Klatt (see p. 12), although it is true that these authors worked with a different kind of speech material. The fact that in connected speech we found small differences supports the claim that there is a universal tendency for *some* difference, however, this difference is not as perceptually relevant as to be the primary marker of voicing of the following obstruent in Czech speakers of English.

Because the original hypothesis was disproved, the author decided to carry out a post-hoc probe – an analysis that was beyond the task set in this thesis – which would help to reveal whether a native speaker's pronunciation actually complies with the hypothesis. It was based on only a very small selection of 29 words from a bulletin read by a male British newscaster ("GF"), one of the source speakers; the same text was read by some of our Czech speakers. Out of these words, 14 contained vowels in fortis contexts and 15 in lenis contexts, both long and short, and the sequences were only VC (not the ones with an intervening sonorant). Normalised vowel durations were once again processed through ANOVA and the outcome was interesting: GF's data for fortis and lenis contexts overlapped just as the rest of the "A" speakers' data did (*cf.* Figure 3.7), which can be seen in Figure 3.10:

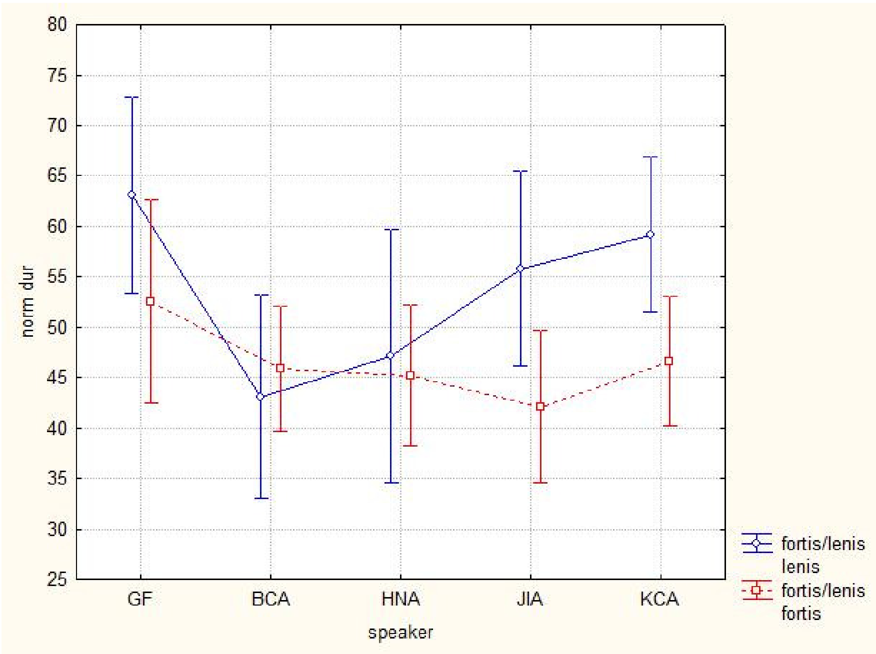


**Figure 3.10.** Mean vowel durations (fortis/lenis contexts) for speakers within category A, with an addition of a native speaker GF (on the left).

Of course, it must be emphasized that the non-existent significance of GF's result stems from the fact that only 29 occurrences were analysed in this probe. This should not be understood as casting doubt on the findings of studies which assert the importance of durational differences for cuing the voicing of the final obstruent in English.

Another notable thing is that the values of vowel durations were generally higher with GF, but otherwise the difference between the mean values for fortis and lenis contexts was roughly the same as KCA's difference. To make the comparison more complex, we used the data illustrated in Figure 3.6 to calculate a similar chart, in that the long and short vowels were distinguished. This brought forth the plain evidence that mainly long vowels are responsible for the overall differences of the mean values of vowel duration in fortis and lenis contexts. In the case of short vowels, the durations did not make any clear-cut distinction possible between fortis/lenis, not even for speaker GF, as one can read from Figure 3.11:

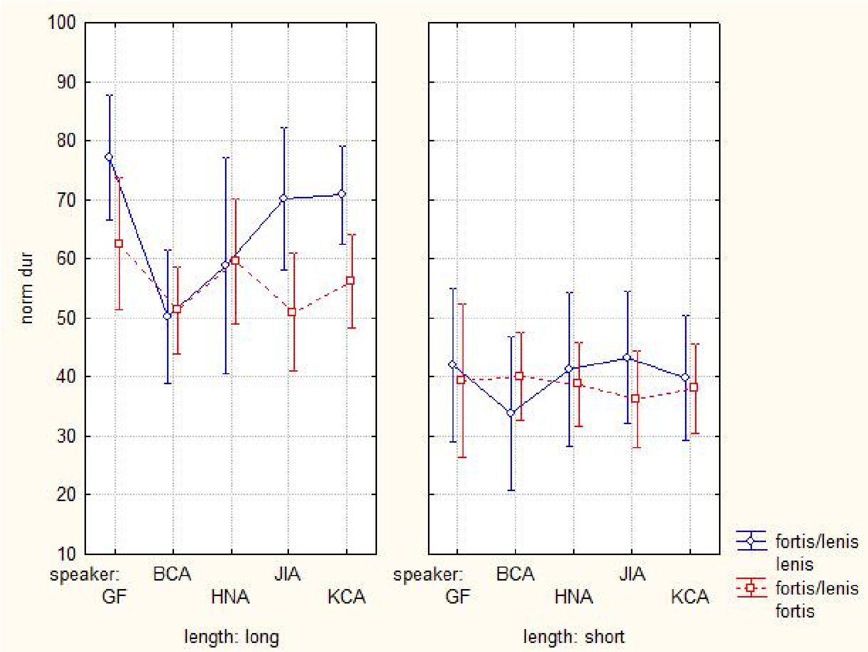


**Figure 3.11.** Mean vowel durations for individual speakers from category A, with the addition of the native speaker GF, taking into account vowel length (left: long vowels, right: short vowels).

Naturally, the fact that the original hypothesis was not confirmed calls for an explanation. One such theory involves speech rate. One feature the "A" speakers and speaker GF had in common was how fluent and relatively fast they were in their readings. It could be suggested that when speech rate exceeds a certain level, the effect of pre-fortis shortening becomes indiscernible. A fact supporting this theory is speaker MILA's undeniable care in pronunciation issuing into a high significance of results. However, this study does not suggest any exact number of syllables per second which could serve as a threshold beyond which the study of the named effects is likely to yield worse results. Besides, giving such a number would border on implying that the pre-fortis shortening phenomenon is hardly traceable in connected speech and therefore cannot be counted among perceptual cues of voicing.

Another explanation this thesis puts forth is that perhaps speakers in category A substitute for the nearly missing pre-fortis shortening by true physical realization of the voicing of the final segment. In other words, a speaker from category A may pronounce “hiss” and “his” with similar vowel duration, but may realize the final segment in “his” with full voicing. Such clear-cut distinction could contribute to him/her being classified as speaker A. On the other hand, speakers from category C would fail to distinguish between “hiss” and “his” in both the temporal domain and in the area of physical voicing. However, this was not the subject of this thesis and remains a possible direction for future study. This could be examined using the voicing profile, as demonstrated by Möbius (2004).

One last factor in favour of our somewhat contrary results is that during the acquisition of English as a second language, Czech pupils and students do not get any explicit instruction as to how to make differences between vowel durations in fortis and lenis contexts. This could indeed be claimed to be a truly fine phonetic detail. It is true that in order to improve the students’ pronunciation, teachers will warn them not to pronounce “swimming” as [swɪmɪŋ**k**] and “there” as [**d**eːr], but pre-fortis shortening is such a delicate phenomenon that the speaker either has to acquire the skill by a good ear or develop an alternative technique of marking voicing, such as described in the preceding paragraph. Despite the fact that the temporal patterning of speech on the subsegmental level probably resists conscious control, and our data indeed fail to differentiate between the three categories of speakers, various levels of temporal patterning are, and will remain an interesting area of phonetic investigations.

# 4 Conclusion

The present thesis departed from a broad overview of various factors and features collected under an umbrella title “the temporal organization of speech”. In the theoretical part, a number of authors were cited whose phonetic research had some implications for the temporal patterning of speech. It was evident from the way the theoretical part unfolded that speech production happens in stages.

First, there is a mental plan in the mind of the speaker, which is an amalgamation of his communicative intentions and restrictions imposed by the rules of the language. In this stage, a majority of the temporal characteristics of the prepared speech is decided, therefore it forms a substantial portion of the theoretical part. It is this stage that contains the choice of marking the voicing of an obstruent by varying the duration of a preceding vowel.

After the first stage, the actual physical speech production follows, in which the temporal pattern is little modified, either by the articulators’ mobility or extraneous conditions (e.g. noise or a special type of the speech addressee). This stage was only noted in passing in various places, where the variation in temporal patterning was to be explained in terms of physical limitations. Apart from that, the acoustic properties of the speech signal played a great role in the empirical part, which was wholly based upon how exactly the speakers produced their pre-fortis vowels in connected speech.

The third stage involves speech perception, a domain that is least accessible to the empirical tools of present-day phonetics, because it is subjective by definition. Perception was relevant for this thesis in that it was decisive for placing the speakers in categories, for identifying a speaker as having a strong Czech accent, for example.

Having introduced the reader to various speech components and their temporal properties, the thesis proceeded to a study comparing the durations of pre-fortis and pre-lenis vowels within the individual categories of speakers, supposing the difference would be greatest in category A (near native-like pronunciation). However, this expectation was not confirmed. The fact that some pre-fortis shortening was detected at all confirmed Chen's assertion that it is, to some extent, common to all languages. The question why category A speakers did not demonstrate a higher tendency for pre-fortis shortening could be answered by suggesting they used other means of marking the voicing of the obstruent. They may have increased the degree of the actual physical voicing, or it may be true that when a certain speech rate is exceeded, the tendency for pre-fortis shortening generally drops cross-linguistically.

Testing the validity of these suggestions could be the matter of future research. This thesis may serve as a modest inspiration for a more detailed inspection in the area of connected speech and temporal patterning.

# Bibliography


Banse, R. and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70, pp. 614-636.

Boersma, P., Weenink, D. (2012). *Praat: doing phonetics by computer* [Computer program]. Version 5.3.14, retrieved on 28 April 2012 from <http://www.praat.org>.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22, pp. 129-159.

Crystal, T. H. and House, A. S. (1988a). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America*, 83, pp. 1553-1573.

Crystal, T. H. and House, A. S. (1988b). Segmental durations in connected-speech signals: Syllabic stress. *Journal of the Acoustical Society of America*, 83, pp. 1574-1585.

Giegerich, H. J. (1992). *English Phonology: An Introduction*. Cambridge: CUP.

Fischer-Jorgensen, E. (1964). Sound duration and place of articulation. *Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft*, 17, pp. 175–207.

Huggins, A. W. F. (1972). Just noticeable differences for segment duration in natural speech. *Journal of the Acoustical Society of America*, 51, pp. 1270-1278.

Jakobson, R. and Halle, M. (1962). Tenseness and laxness. In: R. Jakobson, *Selected Writings I: Phonological Studies*, pp. 550–555. The Hague: Mouton and Co.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59, pp. 1208-1221.

Krakow, R. A. (1999). Physiological organization of syllables: a review. *Journal of Phonetics*, 27, pp. 23-54.

Machač, P. and Skarnitzl, R. (2007). Temporal compensation in Czech? In: *Proceedings of the 16th ICPhS*, pp. 537-540. Saarbrücken: Organizing Committee.

Machač, P. and Skarnitzl, R. (2009). *Fonetická segmentace hlásek*. Praha: Epocha.

Möbius, B. (2004). Corpus-based investigations on the phonetics of consonant voicing. *Folia Linguistica*, 38, pp. 5-26.

Nolan, F. (1983). *Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.

Noteboom, S. (1999). The Prosody of Speech: Melody and Rhythm. In: Hardcastle, W. J. and Laver, J. (Eds.), *The Handbook of Phonetic Sciences*, pp. 640-673, Oxford: Blackwell.

Skarnitzl, R. (2011). *Znělostní kontrast nejen v češtině.* Praha: Epocha.

Skarnitzl, R., Volín, J. and Drenková, L. (2005). Tangibility of Foreign Accents in Speech: the Case of Czech English. In: A.Grmelová, L.Dušková & M.Farrell (Eds.) *2nd Prague Conference on Linguistics and Literary Studies Proceedings*, pp. 11-20. Praha: UK PedF.

Volín, J. (2010). On the significance of the temporal structuring of speech. In: M. Malá & P. Šaldová (Eds.) *...for thy speech bewrayeth thee* (A Festschrift for Libuše Dušková), pp. 289-305. Praha: UK FF.

Volín, J. and Skarnitzl, R. (2007). Temporal downtrends in Czech read speech. In: *Proceedings of Interspeech 2007*, pp. 442-445. Antwerpen: ISCA.

Wells, J. C. (2008). *Longman Pronunciation Dictionary*. Harlow: Pearson Longman.

Yuan, J. and Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics '08.*